Feasibility of Identifying Eating Moments from First-Person Images Leveraging Human Computation

Edison Thomaz, Aman Parnami, Irfan Essa, Gregory D. Abowd

School of Interactive Computing
Georgia Institute of Technology
Atlanta, Georgia, USA

ABSTRACT

There is widespread agreement in the medical research community that more effective mechanisms for dietary assessment and food journaling are needed to fight back against obesity and other nutrition-related diseases. However, it is presently not possible to automatically capture and objectively assess an individual's eating behavior. Currently used dietary assessment and journaling approaches have several limitations; they pose a significant burden on individuals and are often not detailed or accurate enough. In this paper, we describe an approach where we leverage human computation to identify eating moments in first-person point-of-view images taken with wearable cameras. Recognizing eating moments is a key first step both in terms of automating dietary assessment and building systems that help individuals reflect on their diet. In a feasibility study with 5 participants over 3 days, where 17,575 images were collected in total, our method was able to recognize eating moments with 89.68% accuracy.

Author Keywords

Health; Diet; Lifestyle ; Human Computation; Mechanical Turk; Crowdsourcing; Wearable; Egocentric Photos

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

General Terms

Experimentation, Human Factors

INTRODUCTION

The problem of finding out what people eat has been of interest to researchers and individuals for many decades. Formally, two types of methods have been used by researchers to compile dietary information, dietary recalls and records; and food frequency questionnaires. Dietary recalls consist of asking individuals to remember exactly what they ate over a 24h period, while food records require individuals to document what they consume *in situ*, using either a logbook or a mobile phone application, such as MyFitnessPal or MealSnap. Food questionnaires are different in that individuals answer general questions about their diet; questionnaire responses tend to be less detailed and not specific to any particular meal.

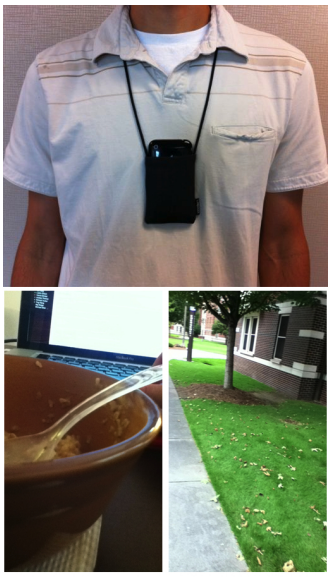

Figure 1. We implemented an application on a standard mobile phone to passively capture first-person point-of-view images.

Much of what we know about the link between diet, health and nutrition we owe to these two classes of methods. However, these techniques suffer from several limitations; they are often tedious for participants, are prone to recall and response bias, are not well suited for longitudinal studies, are plagued by measurement error, and in many cases do not provide detailed enough information to help researchers answer specific questions (e.g. the link between diet and disease) [8, 15].

A recently introduced approach to dietary monitoring involves using wearable cameras such as the eButton [3] and SenseCam [7] to document people's eating behaviors. A head or chest-mounted camera is configured to take first-person point-of-view photos automatically throughout the day (e.g. every 30 seconds), and the resulting snapshots capture people performing a wide range of everyday activities, from socializing with friends to having meals with family members. This

SenseCam 2013 San Diego, USA

technique is particularly promising because it is completely passive; it does not require individuals to do any extra work. Moreover, the images reflect people's eating activities and the surrounding context of those activities truthfully.

However, one of the major challenges of this technique is that only a small portion of the total number of automatically-captured images from a wearable camera depicts an eating activity. Therefore, before these images can be examined from an nutritional perspective or saved in a food journal, it is necessary to devise a mechanism to sift through thousands of first-person point-of-view images and discover the ones that pertain to eating. The sheer volume of images generated per day makes it impractical to annotate them manually, and despite significant progress in the field of computer vision over the years, it remains a challenge to automatically recognize activities in images taken in real world settings.

Over the last few years, human computation has emerged as a viable way to tackle problems that can't be presently solved by computers. Although human computation has been validated as a technique for image labeling [24, 25, 21, 19], identifying health-specific activities in photos through crowdsourcing techniques has not been explored with much depth. In this work, we show how human computation can be applied towards identifying eating moments in first-person point-of-view images. Recognizing eating moments is a key first step both in terms of automating dietary assessment and building systems that help individuals reflect on their diet. Based on a feasibility study with 5 participants over a period of three days, we demonstrate how our system was able to recognize eating moments in real-world settings with 89.68% accuracy.

RELATED WORK

When it comes to inferring eating habits, a number of automatic dietary monitoring approaches have been attempted starting in 1985, when Stellar and Shrager presented an oral sensor to measure chews and swallows during a meal [22]. Sounds from the user's mouth and on-body sensing approaches have been suggested since then to detect when and what individuals are eating [1]. A key finding from this body of research is that no single sensor can capture all dimensions of eating behavior. A different method was tried by Mankoff et al., who relied on shopping receipts to track the nutritional content of foods eaten [13].

Recently, the idea of directly observing individuals from egocentric cameras for overall lifestyle evaluation has been gaining appeal. One of the first cameras used in this context was SenseCam, a lightweight digital camera worn around the neck that passively captures first-person point of view images and sensor readings at regular intervals throughout the day [7]. One of the most unique characteristics of SenseCam is that it doesn't require wearers to perform any action, since images are taken completely automatically. Since its introduction, the SenseCam device has enabled a wide range of applications. Kelly et al. investigated the potential of SenseCam to infer travel research, and in particular evaluate modes and volumes of active versus sedentary travel [9]. Byrne et al. explored SenseCam as a collector of observational data and found it to be complementary to traditional methods. Among other findings, they reported that the passive nature of SenseCam is particularly well-suited for task observations since it doesn't intrude into people's environment [5].

Bai et al. developed a wearable computer called eButton with the goal of "evaluating the human lifestyle" [3]. Similar to the SenseCam in terms of functionality and capabilities, eButton was designed to be worn like a chest button instead of around the neck with a lanyard. It houses a CPU, storage components, a wide-angle digital camera module, and an array of sensors in a small form factor. Sun et al. suggested the use of the eButton for objective dietary assessment [23], and Zhang et al. implemented an activity recognition system from video segments captured with the eButton [26].

Recently, Liu et al. developed a food logging application based on the capture of audio and first-person point-of-view images [12]. The system processes all incoming sounds in real time through a head-mounted microphone and a classifier identifies when chewing is taking place, prompting a wearable camera to capture a video of the eating activity. The authors validated the technical feasibility of their method with a small user study, so it is unclear how their system performs in real world settings. However, applying opportunistic sensors such as microphones towards the problem of eating behavior recognition is a promising direction that we plan to combine with human computation in the future.

Image-Diet Day is another system that automatically captures first-person images [2]. Fourteen participants wore the mobile phone-based device during eating periods for three days and the captured images assisted participants in completing a 24-hour recall procedure. In terms of their value for recall, the images were regarded as helpful, but participants did report technical and perception issues wearing the phone camera device.

Although first-person point-of-view images offer a viable alternative to direct observation, a fundamental problem remains. All captured images must be manually coded for lifestyle indicators, and even with supporting tools such as ImageScape [17] and Image-Diet Day [2], the process tends to be tedious and time-consuming. To address this challenge, we apply human computation through Amazon's Mechanical Turk for this task. Crowdsourcing has matured in the last five years to become an attractive approach to researchers in many fields, including nutritional analysis and activity recognition [16]. One common way to leverage human computation is to use the crowd to provide training data for machine learning classifiers [20]. Even though we see the merits of this approach, in this paper we were particularly interested in whether a classifier could be built using human computation alone.

METHOD

In this paper, we describe a methodology for recognizing eating moments from thousands of first-person point-of-view images by leveraging human computation. The method is comprised of 3 stages, where images are first collected and filtered for privacy protection, formatted into temporal groups,

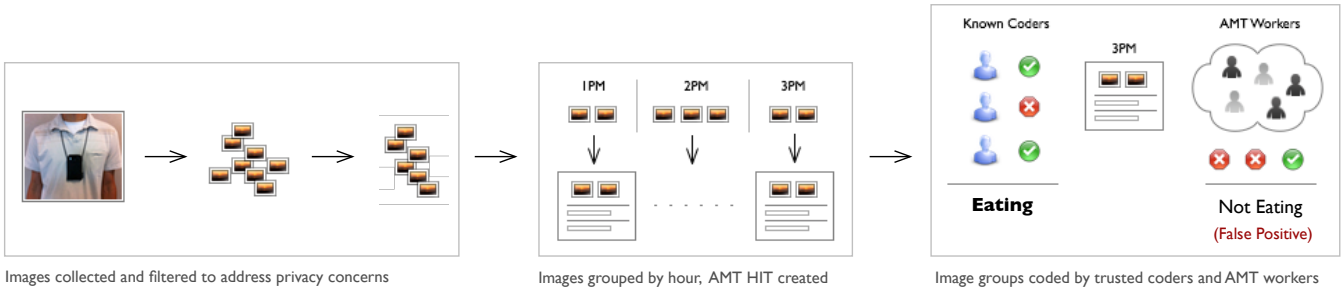

Figure 2. The pipeline for recognizing eating moments from first-person images leveraging human computation and evaluating the performance of the system. It is comprised of 3 stages, where images are first collected and filtered for privacy protection, formatted into temporal groups as a web-based user interface, and finally presented to a group of trusted and human computation workers.

and finally presented to a group of trusted and human computation workers (Figure 2).

Collecting First-Person POV Images

Researchers have used a number of tools for capturing first-person images in the past, such as SenseCam. Because we were interested in using mobile phones for this task and ran into performance issues when testing existing applications that promise this functionality (e.g. Lifelapse), we decided to implement our own app. The additional motivation for having our own implementation was that it could serve as a platform for future experiments and prototypes, as we continue this line of research.

We designed an iPhone application that takes photos automatically every 30 seconds using either the front or back phone camera. People wear the phone as a pendant around the neck with its back-camera facing forward, as shown in Figure 1. All images are saved on the device itself and immediately visible through the built-in "Photos" application. The application is optimized to conserve battery life; it doesn't provide any user interface when running, except for displaying a gray logo on an otherwise completely black background. The only feedback people get from the application is the system's default image snapshot sound effect whenever a picture is taken. If people choose to suppress or minimize this sound effect, they can mute the phone or turn down the volume.

By turning off certain features of the phone, such as Wifi and Bluetooth, and setting the brightness of the screen to its lowest level, we were able to obtain more than 10 hours of battery life on different iPhone models (iPhone 4S, iPhone 4 and iPhone 3G), all running the most recent version of the iOS supported by each device (iOS 6.0.1 and iOS 4.3) at the time of the study.

Excluding Images for Privacy Protection

First-person point-of-view images captured every 30-seconds might depict a day in an individual's life with an unprecedented level of detail. But there is a good chance that these images also reflect aspects of one's life that might be embarrassing or compromising. Therefore, an important step of our method is the exclusion of images that pose a privacy threat to the individuals wearing the camera and to individuals who, knowingly or not, are captured in the images. After transferring all images from the phone to a computer, participants are given the opportunity to review all photos taken by their device and delete any images they would not like to share. Additionally, we review the images and delete any photo that either captured other individuals, or that could reveal sensitive information of the individual who wore the camera. We were required to put these privacy measures in place by our Institutional Review Board (IRB).

Coding Images in AMT

In our method, the task of recognizing eating moments in thousands of first-person point-of-view images is performed by human computation coders. The human computation platform we chose to use was Amazon's Mechanical Turk (AMT). It is described as a "a marketplace for work that requires human intelligence." It exists on the premise that a large number of tasks that computers aren't good at, such as identifying objects in photographs, can be easily carried out by people. Through Mechanical Turk, companies or individuals (called "requesters"), post well-defined tasks ("human intelligence tasks" or HITs) that are matched with, and executed by "workers". Workers signup on the site to perform HITs in exchange for rewards, which range from $0.01 to $1. Requesters can specify a number of parameters for HITs, such as the number of workers that are allowed to perform the task, the qualification of those workers, and the reward amount for tasks completed. Workers are paid only after HITs have been completed and approved by requesters.

Generating HITs

We created a human-intelligence task on AMT that asked workers to examine a group of photos and indicate whether any photo showed an eating activity. If positive, we asked workers for additional information (i.e. meal location and type). The images were grouped by hour, and formatted into a web-based mosaic-like interface (Figure 4). In order to fit a large number of images on the grid, the images are reduced in size, which lowers the amount of activity detail that can be seen. To counter the effect of smaller image sizes, we implemented a script that enlarges the photo underneath the cursor, on hover.

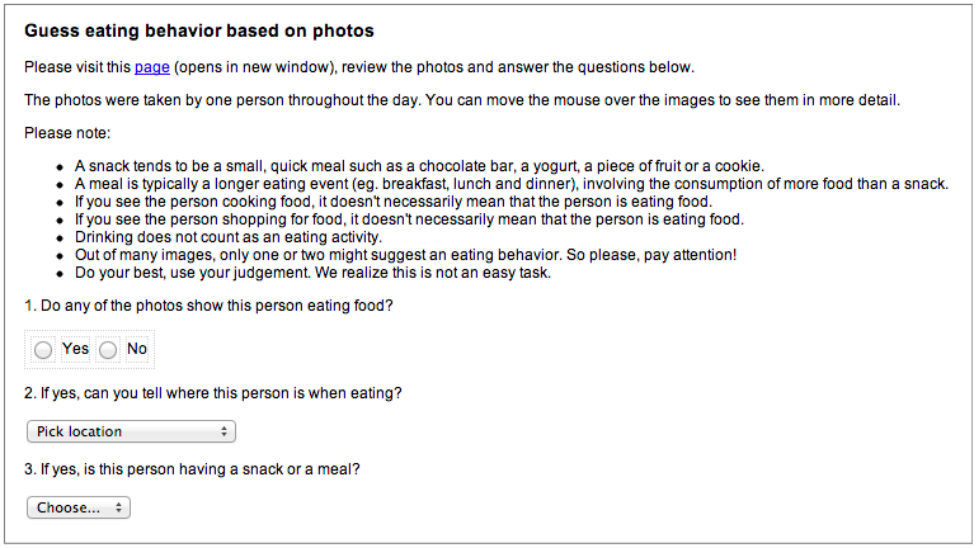

Figure 3. The layout of the human intelligence task (HIT) posted at Amazon's Mechanical Turk for our study. We included a set of guidelines to help workers perform the task successfully. The choices for meal location were: at home, at work or school, at a fast-food restaurant, at a sit- down restaurant, in the car, somewhere else. The choices for meal type were: meal, snack.

Assigning HITs

Once a HIT was created, it had to be assigned to workers. On AMT, it is possible to specify exactly how workers are matched to tasks. To improve the validity of workers' results, we assigned each HIT to three unique workers, and coalesced their votes on each question by taking a majority vote. With this method, depending on the number of workers and valid answers per question (e.g. for meal location), there is a possibility that a majority vote might not be obtained. If and when this condition occurs, the HIT is resubmitted until a majority vote is reached. A completed HIT assignment consisted of the answers to the three questions, the photo group examined, and an identifier for the workers who completed the task.

EVALUATION

We conducted a feasibility study with a non-random convenience sample of participants (n = 5) over 3 days. The only requirement for being in the study was familiarity with the basic operations of a smartphone device. There were 3 females and 2 males, and they ranged in age from 23 to 35 years old and were either graduate students or research scientists at our university. With the exception of one married participant, all other participants were single and either lived alone or with roommates.

Participants were provided with a smartphone preloaded with the custom application, and right before putting on the device, participants were told to verify that the application was running. Participants were instructed to wear the device as much as possible, ideally from the moment they woke up until when they went to sleep. We realized that it would be impractical for subjects to wear the smartphone continuously for hours at a time, so we gave them complete latitude to turn the device off, or take it off if they wanted to or needed to. Due to limited battery life, participants were asked to recharge the device every night.

On average, each participant provided us with 3,509 photos. The image exclusion step where participants reviewed their own images lasted about 15 minutes per participant and led to the removal of up to 200 images. Going through the remaining images and deleting photos that included secondary participants took us at least 45 minutes per subject, and resulted in the deletion of an additional 700 images on average.

One important aspect of Mechanical Turk is that it makes it possible to select workers based on a number of qualifications. Some workers, who have been identified by Amazon as proficient at categorization tasks, are referred to as 'master' workers, and it costs more to recruit them. We hypothesized that our results would be significantly affected by workers' level of qualifications. Therefore, we created identical tasks for categorization masters and regular workers and compared their results. We rewarded all workers $0.15 per assignment and, for regular workers, we indicated that they should have a HIT approval rate greater than 98%.

RESULTS

To assess the performance of Mechanical Turk workers at recognizing eating activities in photos, we had to estimate a measure of ground truth for the image data collected. This was accomplished by having three trusted coders (two authors of the paper and one graduate student) answer the three questions posed in the AMT tasks for each one of the photo groups. The trusted coders used the same web-based interface to examine and browse images as the AMT workers, and their inter-rater reliability was calculated to be 0.65 (Fleiss' kappa).

Table 1 shows how AMT workers performed at identifying eating activities in participants' photos in relation to the estimated ground truth. We calculated recognition accuracy, precision and recall for each participant and across all participants. The results are broken down by worker type to highlight the performance impact of hiring master versus regular

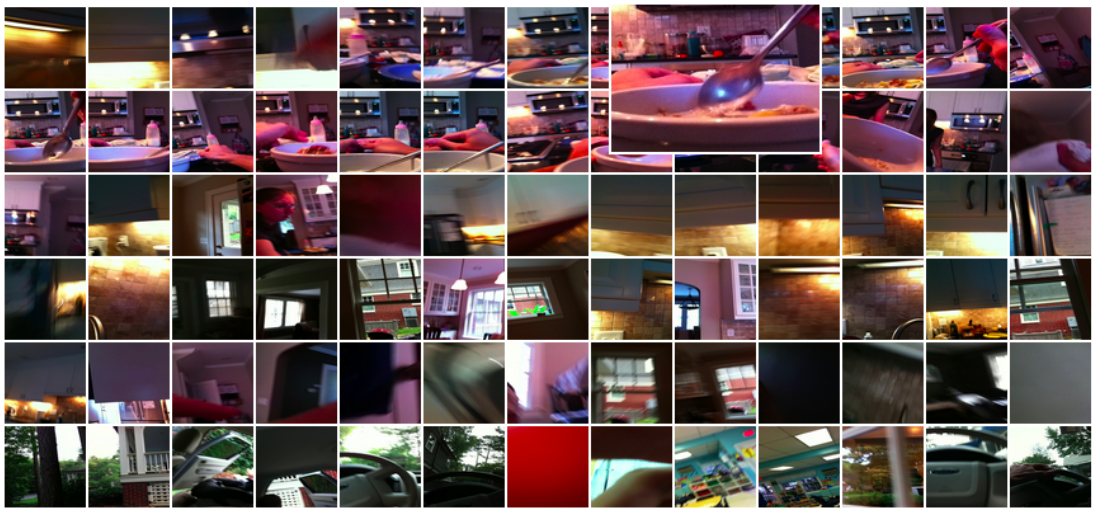

Figure 4. The image grid interface was designed to help Amazon's Mechanical Turk workers browse a large number of photos more efficiently. Hovering the cursor over an images expands it so that it can be examined in more detail, as shown in the middle of the first row.

workers on AMT. As we expected, we saw improved results across all measures when the tasks were assigned to master workers, with overall eating behavior recognition accuracy reaching 89.68% accuracy in the best case scenario. With master workers, overall precision was 86.11% and overall recall was 63.26%.

Inferring meal type and location from first-person point-of-view images is desirable since it might provide additional information that is valuable from a health perspective. However, achieving this from images alone proved to be challenging. Only 19% of meal locations and 24% of meal types were correctly recognized. However, as will be discussed in the next section, these numbers bear little practical significance since meal location can be often obtained through other means in real-world applications (e.g. GPS), and meal type is open to interpretation based on time of day and other factors.

DISCUSSION

One of the most salient results from the evaluation was the low overall recall of AMT master workers (63.26%), indicating that they missed many instances of eating activities. Since each photo group contained upwards of 50 images, it is reasonable that a human might miss important details in the images when constrained by time. This was validated when we confirmed that recall was worse when only one or two photos in a group showed participants eating. This often occurred when the food eaten was consumed quickly, within a minute or two, resulting in the eating behavior being captured in only a small number of photos. We found this to be the case with at least one of the participants, who replaced meals with energy bars.

Overall precision (86.11%) was much closer to overall accuracy for master workers. There were many photos where participants were clearly around food items, such as when shopping for food, in line at a cafe or cooking at home. In most of these cases, one could be easily led to believe that eating was also taking place. This was a common source of false positives in our data. A particularly noticeable result was the disparity in the overall precision measure between regular and master workers. Our results provide evidence that master workers are indeed better at categorization tasks than regular workers, as Amazon claims. This justifies the higher cost paid to AMT to recruit master workers. Overall, for the reasons mentioned above, recognizing eating moments from first-person point-of-view images proved to be a difficult task. This had a direct effect on precision, recall and explains the relatively low agreement reliability amongst coders.

It is important to note that our results only refer to eating activities that were photographed by participants' cameras. Some eating activities might not have been captured. However, given the perspective from which the photos were captured, we feel very strongly that the largest majority of our participants' eating activities was documented.

Meal Location and Type

An individual's location can be often obtained from sensors in mobile phones and other wearable devices. Since there are circumstances when a location sensor is not present or can't be used (e.g. to preserve battery life), we felt that it would be valuable to understand the extent to which meal location could be inferred from images alone. Upon analysis, we were able to attribute the low recognition rates for meal location to two factors. Firstly, because participants wore a phone as a pendant around the neck, all photos were taken at chest-level, pointing directly forward. When participants were sitting at a table and eating, the field of view of the camera was often obstructed by objects in the scene (e.g. body parts, table, chairs, dish-ware, food). This made it difficult to examine the background of the photos and determine participants' whereabouts. We suspect that this issue would have been greatly minimized with the use of a wide-angle camera lens. Secondly, to protect the privacy of secondary participants, we had to discard all photos showing people other than study participants. More often than not, eating is a social activity, with people congregating around a physical space, therefore

Participant	Worker Type	TP	FP	TN	FN	Precision	Recall	Accuracy
P1	regular	5	0	33	9	100%	35.71%	80.85%
	master	10	0	33	4	100%	71.42%	91.48%
P2	regular	1	2	59	10	33.34%	9.09%	83.34%
	master	6	1	60	5	85.71%	54.54%	91.67%
P3	regular	1	1	24	7	50%	12.5%	75.75%
	master	5	0	25	3	100%	62.5%	90.90%
P4	regular	2	2	25	8	50%	20%	72.97%
	master	7	3	24	3	70%	70%	83.78%
P5	regular	1	0	28	5	100%	16.67%	85.29%
	master	3	1	27	3	75%	50%	88.23%
All	**regular**	**10**	**5**	**169**	**39**	**66.67%**	**20.4%**	**80.26%**
	master	**31**	**5**	**169**	**18**	**86.11%**	**63.26%**	**89.68%**

Table 1. Individual and aggregate performance measures showing how well our system was able to identify eating moments from first-person point of view images and human computation. The TP, FP, TN and FN abbreviations refer to true positive, false positive, true negative, and false negative results, respectively.

many of the deleted photos provided rich contextual information about the meal, such as where it took place and with whom. Without these deleted images, it became significantly harder to determine the physical context of the meal.

In terms of meal type, there is a significant amount of ambiguity in what one refers to as a snack or as a meal. Given a photo of a participant eating an energy bar, it is unclear if it should be categorized as a snack or a meal (e.g. lunch). Time of day could be used to help with this differentiation, but ultimately it is a matter of personal interpretation. This interpretive flexibility was reflected in the results for meal type, since our methodology for measuring performance was based on response agreement amongst trusted coders and AMT workers.

Multiple Eating Activities in Photo Group

In our experiment, each photo group included all images captured within a 1 hour interval per participant. We never saw more than one eating activity per photo group. If there had been multiple eating activities within the hour, the exact activity AMT workers based their answers on would have been ambiguous. Spreading all captured photos into more photo groups, each with an interval window of 15 or 30 minutes, would have been a way to address this issue. As previously mentioned, this is an area we plan to explore in future work since we expect that a shorter window might also improve the workers' ability to recognize eating moments.

Mechanical Turk Worker Qualifications

Although the human computation approach offers advantages if compared to a computer vision technique in estimating eating moments from real-world everyday images, it has limitations of its own. One of the characteristics of the method is that people with a wide range of skills and backgrounds are the ones ultimately accepting and completing tasks [18]. Consequently, there is a certain level of variability and non-determinism in human computation that might be unacceptable in certain applications. A set of workers recruited now is always likely to be different from another set of workers recruited just five minutes later.

For a price, it is possible to benefit from a categorization scheme set by Amazon where certain workers are considered to be more proficient at certain tasks than others. We employed both "categorization masters" and regular workers in our study and could verify that results improved significantly with experts. In our experience, seemingly simple parametric modifications in the HIT can have a dramatic impact on performance. There is a large body of research that corroborates this finding, indicating how various factors, from pricing to qualifications, affect the timeliness and quality of the work performed by workers on Mechanical Turk [11, 14, 21].

Privacy

Privacy arouse as an important element of this work, and privacy-related constraints dictated important aspects of our methodology. One of the challenges that we faced was that the wearable camera setup we used (Figure 1) ended up capturing a large number of photos of non-study participants. These included participants' family members, colleagues, neighbors and many other individuals that participants did not know, such as people who happened to be sharing public transportation with participants, visiting the same coffee shop or eating at the same restaurant. Since these individuals were not in our study, they did not consent to their pictures being taken and reviewed by Amazon Mechanical Turk workers.

In order to approve our research, the IRB requested that we delete all such images, which led to the removal of an av-

erage of 700 photos per participant (20% of the total). Importantly, the elimination of these photos had a detrimental impact on the performance of our system, since so many photos of eating activities included secondary subjects. In some cases, more than 90% of a set of images depicting an eating activity had to be deleted. We have no doubts that we will see an improvement in our performance numbers once privacy-protecting measures (e.g. face detection) are put in place and most, if not all, images are available for coding and analysis.

One argument that might be raised about this work is that the benefits gained by crowdsourcing the identification of eating moments in images might be lost due to the effort involved in having to manually review and delete images for privacy reasons. There are two main reasons why we believe this is a weak argument. First of all, the extent to which images had to be reviewed for privacy reasons was stipulated by the IRB. We were required to adopt a protocol where any image that could potentially identify an individual, defined by the presence of any body part of that individual in a photo, had to be eliminated. Understandably, when it comes to privacy matters, the IRB tends to be conservative and we had to abide by its rules in order to conduct our study. However, due to many reasons (e.g. the popularity of wearable devices such as Google Glass), the principles that guide privacy policy might change in the future. Our numbers represent a lower-bound in terms of performance and a change in privacy policy will likely result in our ability to delete fewer images. Secondly, the growing trend of using wearable cameras and first-person point-of-view images in health research has brought the issue of privacy to the forefront. In future work we plan to adopt emerging strategies and techniques for mitigating privacy concerns within this technological context [10]. We believe this will be an additional factor lowering the manual effort involved in the methodology.

CONTRIBUTIONS

Our long term goal is to develop automated and semi-automated dietary assessment systems to be used in real-world settings. A promising approach towards this goal involves automatically documenting people's eating activities with photos taken with a wearable camera at regular intervals (e.g. every 30 seconds). Recognizing eating moments in photos is a key first step before these photos can be analyzed for nutritional information through a system like PlateMate [16], or added to a food journal. The contribution of this work is to show that human computation can be successfully used for sifting through first-person point-of-view images captured with a wearable camera and identifying eating moments. Although human computation has been validated as a technique for image labeling [24, 25, 21, 19], identifying health-specific activities in photos through crowdsourcing techniques has not been explored with much depth. In a feasibility study with 5 participants over 3 days, where 17,575 images were collected in total, forty nine instances of eating activity were recorded in the photos and identified by AMT workers with 89.68% accuracy.

FUTURE WORK

There are a myriad of opportunities when it comes to extending this work, both in depth and breadth. In terms of methodology, the strategy of labeling images through majority vote is the only crowdsourcing quality control we use. It is certainly an effective one, as it accounts for occasional human errors and variability in human performance [21]. Hara et al. studied the impact of accuracy in majority group size and determined that performance gains diminish significantly as group size grows beyond 5 AMT workers [6]. For cost reasons, we kept majority vote group size to 3 workers in this feasibility study. In the future we plan to put in place additional quality measures as well, such as validation or Find-Fix-Verify [4]. With validation, a set of AMT workers evaluate the classification of images that have already been labeled.

In this paper, we showed how modifying parameters that configure the creation of Mechanical Turk's HITs can have a very substantial impact in the quality of the workers recruited and, consequently, the quality of the recognition job performed. A fruitful area for exploration would be to study in more detail how manipulating the worker qualifications variable affects recognition accuracy, and at what cost. Likewise, it would be desirable to examine the extent to which the frequency images are taken affects eating behavior inference. More images will lead to a more comprehensive visual account of an individual's activities and eating moments, but at the expense of more image analysis and camera battery life. We also don't know the extent to which the use of our mobile phone camera influenced people's eating behaviors.

In terms of tradeoffs, a very significant one exists between privacy and camera placement. When located on an individual's head, a regular camera is much more likely to take photos that are representative of the activity the individual is engaged in. This is because the camera will always be pointing in the direction the head is pointing to. On the other hand, because it is elevated, the camera records much more of the individual's surrounding, which includes other people. We are interested in exploring how to mitigate the privacy challenges that are inherent in this technique, such as by attempting to selectively and temporarily change the capture viewport of the camera.

ACKNOWLEDGMENTS

We would like to thank the Intel Science and Technology Center for Pervasive Computing (ISTC-PC) for supporting this work.

REFERENCES

1. Amft, O., Stäger, M., Lukowicz, P., and Tröster, G. Analysis of chewing sounds for dietary monitoring. In *UbiComp'05: Proceedings of the 7th international conference on Ubiquitous Computing*, Springer-Verlag (Sept. 2005).

2. Arab, L., Estrin, D., Kim, D. H., Burke, J., and Goldman, J. Feasibility testing of an automated image-capture method to aid dietary recall. *European Journal of Clinical Nutrition 65*, 10 (May 2011), 1156–1162.

3. Bai, Y., Li, C., Yue, Y., Jia, W., Li, J., Mao, Z.-H., and Sun, M. Designing a wearable computer for lifestyle evaluation. In *Bioengineering Conference (NEBEC), 2012 38th Annual Northeast* (2012), 93–94.
4. Bernstein, M. S., Little, G., Miller, R. C., Hartmann, B., Ackerman, M. S., Karger, D. R., Crowell, D., and Panovich, K. Soylent: a word processor with a crowd inside. *UIST* (2010), 313–322.
5. Byrne, D., Doherty, A. R., Jones, G. J. F., Smeaton, A. F., Kumpulainen, S., and Järvelin, K. The SenseCam as a tool for task observation. In *Proceedings of the 22nd British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction*, British Computer Society (Sept. 2008).
6. Hara, K., Le, V., and Froehlich, J. Combining crowdsourcing and google street view to identify street-level accessibility problems. In *CHI '13: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM Request Permissions (Apr. 2013).
7. Hodges, S., Williams, L., Berry, E., Izadi, S., Srinivasan, J., Butler, A., Smyth, G., Kapur, N., and Wood, K. SenseCam: a retrospective memory aid. In *UbiComp'06: Proceedings of the 8th international conference on Ubiquitous Computing*, Springer-Verlag (Sept. 2006).
8. Jacobs, D. R. Challenges in research in nutritional epidemiology. *Nutritional Health* (2012), 29–42.
9. Kelly, P., Doherty, A., Berry, E., Hodges, S., Batterham, A. M., and Foster, C. Can we use digital life-log images to investigate active and sedentary travel behaviour? Results from a pilot study. *International Journal of Behavioral Nutrition and Physical Activity 8*, 1 (May 2011), 44.
10. Kelly, P., Marshall, S. J., Badland, H., Kerr, J., Oliver, M., Doherty, A. R., and Foster, C. An ethical framework for automated, wearable cameras in health behavior research. *American journal of preventive medicine 44*, 3 (Mar. 2013), 314–319.
11. Kittur, A., Chi, E. H., and Suh, B. Crowdsourcing user studies with Mechanical Turk. In *CHI '08: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, ACM Request Permissions (Apr. 2008).
12. Liu, J., Johns, E., Atallah, L., Pettitt, C., Lo, B., and Frost, G. An Intelligent Food-Intake Monitoring System Using Wearable Sensors.
13. Mankoff, J., Hsieh, G., Hung, H. C., Lee, S., and Nitao, E. Using Low-Cost Sensing to Support Nutritional Awareness. In *UbiComp '02: Proceedings of the 4th international conference on Ubiquitous Computing*, Springer-Verlag (Sept. 2002).
14. Mason, W., and Watts, D. J. Financial incentives and the "performance of crowds". In *HCOMP '09: Proceedings of the ACM SIGKDD Workshop on Human Computation*, ACM Request Permissions (June 2009).
15. Michels, K. B. Nutritional epidemiology–past, present, future. *International journal of epidemiology 32*, 4 (Aug. 2003), 486–488.
16. Noronha, J., Hysen, E., Zhang, H., and Gajos, K. Z. Platemate: crowdsourcing nutritional analysis from food photographs. *Proceedings of the 24th annual ACM symposium on User interface software and technology* (2011), 1–12.
17. Reddy, S., Parker, A., Hyman, J., Burke, J., Estrin, D., and Hansen, M. Image browsing, processing, and clustering for participatory sensing: lessons from a DietSense prototype. In *EmNets '07: Proceedings of the 4th workshop on Embedded networked sensors*, ACM Request Permissions (June 2007).
18. Ross, J., Irani, L., Silberman, M., Zaldivar, A., and Tomlinson, B. Who are the crowdworkers?: shifting demographics in mechanical turk. *Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems* (2010), 2863–2872.
19. Russell, B. C., Torralba, A., Murphy, K. P., and Freeman, W. T. LabelMe: A Database and Web-Based Tool for Image Annotation. *International Journal of Computer Vision 77*, 1-3 (May 2008).
20. Song, Y. C., Lasecki, W. S., Bigham, J. P., and Kautz, H. Training Activity Recognition Systems Online Using Real-time Crowdsourcing. *Ubicomp '12: Proceedings of the 14th ACM International Conference on Ubiquitous Computing* (2012).
21. Sorokin, A., and Forsyth, D. Utility data annotation with Amazon Mechanical Turk. *Audio, Transactions of the IRE Professional Group on* (June 2008), 1–8.
22. Stellar, E., and Shrager, E. E. Chews and swallows and the microstructure of eating. *The American journal of clinical nutrition 42*, 5 (1985), 973–982.
23. Sun, M., Fernstrom, J. D., Jia, W., Hackworth, S. A., Yao, N., Li, Y., Li, C., Fernstrom, M. H., and Sclabassi, R. J. A wearable electronic system for objective dietary assessment. *Journal of the American Dietetic Association 110*, 1 (2010), 45.
24. von Ahn, L., and Dabbish, L. Labeling images with a computer game. In *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM Request Permissions (Apr. 2004).
25. von Ahn, L., Liu, R., and Blum, M. Peekaboom: a game for locating objects in images. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, ACM Request Permissions (Apr. 2006).
26. Zhang, H., Li, L., Jia, W., Fernstrom, J. D., Sclabassi, R. J., and Sun, M. Recognizing physical activity from ego-motion of a camera. *Proceedings of IEEE EMBS 2010* (2010), 5569–5572.